Assessing the Severity of Color Vision Loss with Implications for Aviation and other Occupational Environments

Marisa Rodriguez-Carmona# (MSc, PhD), Matilda O'Neill-Biba (BSc, MCOptom) & John L. Barbur (MSc, PhD)

Applied Vision Research Centre, The Henry Wellcome Laboratories for Vision Sciences, City University, Northampton Square, London. EC1V 0HB, UK

to whom correspondence should be addressed

Short title: Severity of color vision loss

Word count for Abstract: 250

Word count for Narrative: 5850

Number of References: 23

Number of Tables: 2

Number of Figures: 9

Abstract

Introduction: The Ishihara Test (IT) is arguably the most sensitive and commonly used color vision test within aviation and other occupational environments, but when no errors are allowed ~20% of normal trichromats fail the test. The number of allowed errors varies in different occupations and sometimes within the same environment (such as aviation) in order to reflect the difficulties of the color-related tasks. The implicit assumption is that the plates can be ranked in order of difficulty. The principal aim of this study was to investigate whether appropriate ‘weights’ can be attached to each IT plate to reflect the likelihood of producing a correct response. A second aim was to justify the use of color thresholds for quantifying the loss of red-green (RG) and yellow-blue (YB) chromatic sensitivity.

Methods: We investigated 742 subjects (236 normals, 340 deutans and 166 protans) using the first 25 plates of the 38-plate IT and measured RG chromatic sensitivity using the Color Assessment and Diagnosis (CAD) test. The IT error scores provided plate-specific ‘weights’ which were used to calculate a Severity Index (SI) of color vision loss for each subject.

Results: Error scores, SI values and CAD thresholds were measured and compared in each of the three subject groups.

Conclusions: Color thresholds can provide a good measure of the severity of both RG and YB color vision loss. Neither the number of IT plates failed nor the SI value computed in this way can be used to determine reliably the severity of color vision loss.

Keywords: Ishihara pseudoisochromatic plates; CAD test; color deficiency; cone contrasts

Introduction

The Ishihara pseudoisochromatic plates and other similar color screening tests make use of static luminance and chromatic contrast masking to isolate the use of color signals and to reveal loss of red-green (RG) chromatic sensitivity. The Ishihara Test (IT) is the most extensively used color vision test and has been shown to be very sensitive for RG color deficiency (5,11,20), but this is achieved at the expense of specificity. The IT was first published in 1917 and since then it has been reprinted in many different editions. Currently three versions are available: the full version containing 38 plates, the abbreviated version with 24 plates and the concise version containing 14 plates. The full version has been recommended for routine clinical use (7), albeit the abbreviated and concise versions being subsets of the full version. In the full version, the first 25 plates contain numerals (single or double digit numbers) which have to be named verbally, whereas the remaining plates have pathway designs intended for examination of non-verbal subjects and are rarely used. The IT employs a range of designs that can be divided into five categories (17):

1. Introductory or demonstration: plate 1 – read by all subjects including congenital color deficients.
2. Transformation or confusion: plates 2-9 – normal trichromats and subjects with color vision deficiency read different numbers.
3. Vanishing: plates 10-17 – normal trichromats read a single number, while color vision deficients usually do not see any number.
4. Hidden digit: plates 18-21 – color deficients can read a number, whilst normal trichromats do not.

5. Classification: plates 22-25 – designed to distinguish between protan and deutan subjects.

Mistakes on the pseudoisochromatic plates have been described as either typical or non-typical (16). Errors made by color vision deficient subjects are considered to be typical, whilst non-typical errors are usually made by normal trichromats. Although errors are in general made as a result of either weak or varying chromatic signals and hence poorly defined and / or confusing chromatic contours, other individual related factors such as age, gender, observation time and even school achievement have been put forward as important, particularly when non-typical errors are involved. Digit confusion, as an example of a non-typical error, has been described previously and such confusions are more likely to occur when weak color signals are involved (7,9,12). Birch and McKeever have described non-typical errors made by normals as ‘‘misreadings” - as a partial filling up of serif design of the digits of the IT, for example a ‘5’ may be interpreted as a ‘6’, or ‘3’ as ‘8’ (7). Although such factors may well contribute to a ‘misreading’, these factors apply equally well to subjects with congenital color deficiency in the context of a numeral defined by weaker and often different chromatic signals. In general, the weaker the perceived chromatic signal in the test plate, the higher the probability of confusions or misreadings.

The IT is used widely and has very high sensitivity for detecting subjects with mild congenital deficiency when no errors are allowed, but the outcome is not specific since a large percentage of subjects with normal color vision can also make some errors when reading the IT plates (9,12,15). Due to the simplicity, availability and low cost, the IT is relatively easy to learn, making use of cues other than color, i.e., learning how to use plate-specific pattern features located at easily identifiable positions along vertical and horizontal lines that pass through the centre of each plate. Tracing the wavy lines on the last 12 plates is more difficult to learn using pattern features, but these plates are rarely used in certification. The use of other cues to improve one's IT score may also be possible given the extremely high motivation some applicants have to pass the test, i.e., rehearsing the correct response through the use of achromatic cues in advance of the test, depending on the pattern of particular plates perceived by the observer. There may also be significant differences amongst subjects with deutan- and protan-like deficiency in the way they perform on the IT plates, particularly when only mild loss of chromatic sensitivity is involved. Since dichromats and subjects with severe loss of RG chromatic sensitivity tend to fail most of the IT plates, it has often become the practice in many occupational environments to use the number of plates failed as a direct measure of the severity of color vision loss. This practice rests on at least two implicit assumptions:

1. The IT plates follow some linear trend when ranked in order of difficulty
2. This order remains unchanged for subjects with deutan and protan deficiency, i.e., the same number of errors reflects similar loss of chromatic sensitivity in subjects with either deutan or protan deficiency

The need to assess accurately the severity of color vision loss has recently become more important simply because many color deficient subjects have been shown to have sufficient, residual RG chromatic sensitivity to perform visually demanding, color-related tasks with the same accuracy as normal trichromats (8). The easy solution to this problem is to allow a small number of errors on the IT plates when screening for 'functionally safe' color vision, depending on the color-related visual demands within a given occupation. Current practices in a number of selected environments are summarized in Table I together with the corresponding percentages of subjects that pass the requirements within each subject group. In commercial aviation, the Joint Aviation Requirements (JAR) approve the use of the IT as the primary screening test. A pass requires the applicant to read correctly the numbers on each of the first 15 plates. If the applicant fails this test a secondary test is carried out, usually an anomaloscope match or a lantern test (14). Secondary testing is an inevitable requirement when using the strict pass/fail criterion on the IT plates and ensures that applicants with normal color vision are likely to pass. The Federal Aviation Administration (FAA) accepts a number of primary tests (no secondary tests), which include the Ishihara test amongst others (10). The FAA guidelines for the use of the IT plates state that the applicant should be certified as safe with *k* or less errors (Table I). This number varies from five or less errors on plates 1-11 on the concise 14-plate edition, six or less errors on plates 1-15 on the abbreviated 24-plate edition and eight or less errors on plates 1-21 on the full 38-plate edition (10). Table I reveals that within the FAA the percentage of applicants that pass depends on the edition of the IT used. According to the FAA criteria, the concise 14-plate edition passes the largest percentage of colour deficient subjects. Overall, more applicants with congenital colour deficiency will pass the FAA when compared to the JAR accepted test protocol. In other occupations, such as the Fire Service in the UK, a pass requires two or less errors on the first 17 plates of the IT 24-plate edition and London Underground Ltd. (UK), before the introduction of CAD based pass / fail limits in 2008, allowed up to three 'errors on specified plates' (esp), i.e., plates 1-17 of the 38-plate edition, which have been defined as "misreadings" (6). In the absence of comparative studies it is difficult to establish the extent to which the number of plates failed on the IT can be taken as a measure of the severity of RG color vision loss.

[Table I approximately here]

The purpose of this study is to examine the extent to which the number of IT plates failed provides a linear measure of the severity of color vision loss and whether the same measure can be applied to both classes of congenital RG color deficiency. We examine specifically whether the various IT plates can be weighted separately to reflect the probability of a correct response for normal trichromats and for subjects with deutan or protan color deficiency. In order to achieve this aim, we also carried out some analysis of threshold measures of RG and YB chromatic sensitivity to establish how these thresholds relate to the corresponding cone contrasts generated and hence the severity of color vision loss. The cone contrast, for each of the three cone classes, is defined as the difference in cone signal between the test and the background expressed as a fraction of the background signal, i.e.

$$C = (S_{test} - S_{bkg}) / S_{bkg}$$

where S_{test} and S_{bkg} represent the signals generated by the test stimulus and the adjacent background, respectively (see Fig. 1).

Methods

Subjects

The 742 subjects that participated in this study were mainly volunteering students and staff from City University as well as occupational applicants referred to City University for advanced color vision assessment. 236 subjects had normal color vision (136 males, 100 females) as confirmed by testing with the Nagel anomaloscope (Type I) and the Color Assessment and Diagnosis (CAD) test. The remaining 506 participants had deficient color vision, of whom 340 had deutan and 166 had protan deficiency. The mean age of the subjects was 31.0 ± 11.7 years, with a median of 28 years. Subjects with best corrected visual acuity less than 6/12 or recorded clinical signs of pathology or ocular abnormalities were excluded from the study. All experiments were conducted in accordance with the tenets of the Declaration of Helsinki (Code of Ethics of the World Medical Association), and the study had the approval of the Research and Ethics Committee of City University.

Assessment of chromatic sensitivity

The first 25 plates of the Ishihara pseudoisochromatic plates (Kanehara & Co. Ltd, Tokyo, Japan, 38-plate edition, 1989) were used with each of the 742 subjects. The Macbeth Easel desk lamp (designed for use with plate tests) was employed, on average the subjects took less than 5 s per plate, the viewing distance was approximately 60 cm, and each error made was recorded. As expected, all subjects read correctly the number on the introductory plate (plate 1). In addition, every subject had his / her color vision assessed using the CAD test. This employs a calibrated visual display and consists of colored stimuli of specified luminance and chromaticity (see Fig. 1) embedded in a background of dynamic luminance contrast (LC) noise

(2,3). The square outline stimulus moves along each of the diagonal directions and the subject’s task is to indicate the direction of motion of the color-defined stimulus. An efficient, four-alternative forced-choice procedure is used to measure the subject’s chromatic detection thresholds along 16 directions in the CIE 1931 – (x,y) chromaticity chart. The stimulus directions were selected to ensure automatic classification of the class of deficiency involved, as well as adequate estimates of both RG and YB thresholds. 16 randomly interleaved staircases are employed and the full test takes approximately 12 min to complete. The median thresholds measured for 330 normal trichromats (age range 14–60; mean = 30 ± 10 years) define the standard normal (SN) CAD observer. The statistical distributions of RG and YB thresholds were also used to calculate the corresponding 2.5% and 97.5% limits (18,19). The SN observer parameters provide an efficient way of assessing the severity of color vision loss, i.e., an observer with a RG threshold of 2 SN CAD units requires twice the color signal strength needed by the standard CAD observer. It is of interest to establish how well the CAD thresholds reflect the severity of color vision loss.

RG and YB thresholds as a measure of the severity of color vision loss

[Fig. 1 here]

Although the CIE (x,y) chart is non-linear, the measured RG and YB thresholds can be predicted accurately for any state of chromatic adaptation and over a large range of luminances based on the cone excitation signals produced by the adapting background field (13). In this respect the diagram is linear in that the change in cone signals at threshold is a constant fraction of the background cone excitation signals for each cone class. The CAD test employs a daylight (D_{65}) background and the color thresholds are measured as a chromatic displacement (CD) away from background chromaticity, towards a specified point on the spectrum locus. An interesting and useful property of the (x,y) – chart is that the cone contrasts generated along a line such as that shown in Fig. 1a increase almost linearly with CD distance for all three cone classes, even up to CD values as large as 20 SN threshold units, Fig. 1e. The steps involved in the computation of cone contrasts are illustrated in Fig. 1. The method assumes only knowledge of the spectral responsivity functions of the cones and can also be applied to any other photoreceptor pigments (e.g., rods or melanopsin). Figure 1a shows the median threshold ellipse computed from measurements of chromatic detection thresholds in 16 directions of color space in 330 normal trichromats using the CAD test. A uniform background of luminance 24cd m^{-2} and chromaticity 0.305, 0.323 was employed. The mean luminance of the colored test stimulus remains the same as that of the surrounding background. The direction of the line (i.e., the grey arrow in section a) is specified as an angle (see Fig. 2) measured with respect to the abscissa in an anti-clockwise direction. The contrast generated by the test stimulus in each class of cone photoreceptor is shown in Fig. 1e as a function of chromatic displacement along the selected line. In order to understand the steps involved in the computation of cone contrast it is best to start with Fig. 1a and to follow the diagrams in a clockwise direction as indicated by the arrows. The spectral radiance of the background is known and for any point on the grey line shown in Fig. 1a one can compute the spectral radiance of the stimulus (as shown for illustration in Fig. 1b). The signals generated by the stimulus and the surrounding background in each cone photoreceptor class are then computed from the spectral radiance data (section b) and the normalised spectral responsivity functions of the cones (section c). The cone contrasts shown in Fig. 2 were computed for stimuli that fall on the median ellipse for normal trichromats shown in Fig. 1a. The direction of CD towards any point on the spectrum locus is determined by its corresponding hue angle (measured with respect to the horizontal axis in an anti-clockwise direction). As an example, the "yellow" region of the spectrum locus corresponds to an angle of ~ 67^o, whilst the blue region corresponds to an angle of 247^o. The arrows in Fig. 2b show clearly the directions of CD that correspond to approximately zero L- and M- cone contrasts: ~ 67^o and 247^o (i.e., the YB-axis or the tritan line) and also the two directions for which the S-cone contrast is zero: 154^o and 334^o (i.e., the RG-axis). When one examines the cone contrast needed at threshold for the median normal trichromat, along the RG axis the median RG threshold requires ~ 0.75% and 0.4% M- and L-cone contrast, respectively. Along the tritan axis the median S-cone contrast at threshold is 8%. YB color discrimination at threshold is therefore significantly less sensitive than RG discrimination and this may reflect the much smaller relative number of S cones in the retina (1). Fig. 3 shows cone contrasts as a function of CD value computed for 67^o and 247^o (i.e., the tritan line) and for 154^o and 334^o (i.e., the RG axis) to illustrate how cone contrasts increase with CD along the YB and RG axes. These directions were selected to reflect the cone contrast changes along the RG and YB axes. The results show that RG and YB thresholds relate almost linearly to the corresponding cone contrasts generated and this suggests that the magnitude of the measured thresholds may provide a good indicator of the severity of color vision loss. There are a number of other interesting observations that emerge from the cone contrasts shown in Fig. 3. The YB directions show greater non-linearity than the RG directions. For thresholds less than about some five SN CAD units, the departure from linearity for both increments and decrements in S-cone contrast is small, but increments in S-cone contrast increase more rapidly than decrements at larger CD values. This observation suggests that under mesopic light adaptation when YB chromatic sensitivity is poor and large CDs are needed, thresholds in the blue direction should be smaller than the corresponding thresholds towards the yellow region of the spectrum locus and this has been confirmed experimentally (22).

[Fig. 2 and 3 approximately here]

When dealing with color vision, it is difficult to relate the severity of color vision loss to parameters such as the number of IT plates the subject fails or the midpoint and range measured in anomaloscope matches (4,23). It is, however, generally accepted that visual performance relates well to luminance contrast and that the loss of contrast sensitivity in the case of achromatic stimuli is a good predictor of loss of visual performance. A new experiment was therefore designed to measure the contrast of an achromatic stimulus that matches the perceived contrast of a colored stimulus in the absence of any photopic or scotopic luminance contrast. We selected CDs along a line in the CIE -1931 chart, specific to the CRT display employed, that correspond to zero rod contrast (i.e., $\theta = 117^{o}$ and 297^{o}). The background had CIE (x,y) chromaticity 0.305, 0.323 and luminance was 10cd

m^{-2}. For each CD value selected we measured the luminance contrast of a spatially identical achromatic stimulus that matched the perceived contrast / conspicuity of the colored stimulus. The results shown in Fig. 4 reveal an almost linear relationship between the strength of the colored signal (defined by its CD value) and its “equivalent” achromatic contrast. These findings suggest that although the strength of color signals remains difficult to define accurately, the threshold CD measured in the CIE-(x,y) chart represents a good indicator of the severity of the subject’s color vision loss. In view of these findings, the measured RG and YB thresholds will be used as an index to describe the severity of color vision loss.

[Fig. 4 approximately here]

Chromatic sensitivity varies almost continuously when measured in subjects with deutan- or protan-like congenital deficiency. The severity of color vision loss varies from complete absence of RG discrimination, in the case of dichromats, to almost normal sensitivity in subjects with thresholds not much larger than 2 SN CAD units. The loss of RG sensitivity (when expressed in SN CAD units) is greater in protanomalous than deuteranomalous subjects (4,8). The most severe color vision losses measured with the CAD test may exceed 20 SN CAD units for RG discrimination when the subject’s thresholds are often limited by the phosphors of the display. The YB thresholds on the other hand show little variation, as expected in the absence of yellow-blue loss or acquired deficiency.

Computation of Severity Index (SI) for the Ishihara Test

It is of interest to establish whether the introduction of a new index that takes into account the plate-specific probability of a correct response provides an improved measure of the severity of color vision loss. We used the Ishihara error scores to establish appropriate 'weights' for each of the 25 plates, separately, for each subject group. A severity index was then computed for each subject by summing up the corresponding 'weights' for the failed plates. The plate-specific 'weight' reflects the probability of a correct response. If a large percentage of subjects within a group make errors on a given plate, the probability of a correct response is small and the plate has a lower 'weight'. Conversely, if a plate is read incorrectly by only a small percentage of subjects within a group, the probability of a correct response is large and hence the plate is given a large 'weight' simply because only the most severe subjects will fail this plate.

Statistical analysis

We define the Severity Index (SI) based on the errors the subject makes as: $SI=\sum_{i=2}^{25} W_i * R_i$, where '$R_i$' is used to indicate the subject's response (R_i=1, indicates an incorrect response and R_i=0, indicates a correct response).

The Weight (W_i) of a plate for each subject group (N, D, P) is simply the plate-specific probability of a correct response and is given by:

$W_i=1-PN_i;\ 1-PD_i;\ 1-PP_i$ where PN_i, PD_i and PP_i is the probability of normals, deutans and protans, respectively, making an error on plate i.

In addition, the weights within each group are multiplied by a constant to ensure that SI values fall within a range of zero to 100. A value of zero corresponds to no errors and indicates 'perfect' color vision, whilst a value of

100 corresponds to the maximum number of errors (i.e., 24) and indicates complete absence of RG color vision. Since the weight is proportional to the probability of a correct response, this can be written as:

$$W_i=k*(1-PE_i),$$

Where, PE_i, represents the measured probability of an incorrect response for plate, i, within each subject group, and

$$k=100/\sum_{2}^{25}(1-PE_i).$$

Results

The probability of a subject making *k* or less errors on the IT plates is shown in Fig. 5 for normal trichromats and for subjects with deutan and protan deficiency. The results show that 80.9% (191) of normal trichromats make no errors on the first 25 plates of the 38-plate version and almost all normals (except for 1) get all 25 plates correct with 3 or less errors. Definition of normal colour vision in terms of the number of errors allowed depends on the edition employed. Normality as per the instruction manual for the 38-plate version is determined if 17 or more plates are read correctly on plates 1-21. 10% of deutan and 1% of protan subjects also make four or less errors. Fig. 5 shows that the probability of making *k* or less errors is much greater for deutan than for protan subjects. If the number of IT plates failed is a valid indicator of the severity of color vision loss then the results of Fig. 5 demonstrate that for the same number of errors made, the severity of color vision loss is much greater in protan than deutan subjects. For example, 29% of deutan subjects make 12 or less errors compared with only 8% of protan subjects. Similarly, 70% of deutan subjects make 20 errors or less compared with only 39% of protan subjects. The majority of protan subjects make at least 21 errors. Note that no subject made errors on the introductory plate, hence the maximum number of possible errors is 24.

[Fig. 5 here]

Figure 6 shows the percentage of plate-specific errors made within each subject group ranked in sequence. Data are shown for normal trichromats (bottom) and for subjects with deutan (top) and protan deficiencies (middle). In the case of deutans, for example, 92% make errors when presented with plate 12, but only 29% fail to read correctly the number on plate 21. This makes plate 21 easier to pass and is therefore less challenging to a deutan subject than plate 12. Consequently, the subjects that fail plate 21 are likely to have more severe loss of RG color vision. In general, a larger number of protan subjects produce errors on every plate when compared to deutans. In particular, plates 21, 5 and 18 are read correctly by over 60% of deutan subjects, whilst plate 21 is the least challenging for protan subjects with 49% correct responses. The first four 'transformation' plates (plates 2-5) and the 'hidden digit' design plates (plates 18 to 21) produce the largest differences in error rates between deutan and protan subjects. Plates 12 and 17 (i.e., vanishing plate designs) present almost the same level of difficulty for both deutan and protan subjects. Interestingly, plates 12 and 17 are also the two most difficult plates for normal trichromats with 11% and 6% error rates, respectively. Within the deutan group, the easiest plate corresponds to an error rate of 29% whilst the most difficult plate to 92%. The equivalent variation is from 52% to 98% for protan subjects and 0% to 11% for normal trichromats. Although the level of difficulty per plate shows less variation within the protan group, the difference in ranking order shown in Fig. 6 suggests that some benefit could be secured by ‘weighting’ each error made by the plate-specific probability of a correct response within each subject group.

[Fig. 6 here]

Table II analyses in greater detail the errors normal trichromats make on the IT plates by summarizing the type and frequency of errors normals made in this study. The numbers shown in brackets represent the number of times that error response was observed within the group. For example, number 3 displayed on plate number 7 is reported as number ‘8’ by five normal trichromats from a total of 236 subjects. Note that five normal subjects read number ‘2’ on plate 19 which is also a response typical in subjects with color deficiency (shown in bold font).

[Table II here]

Figure 7a plots the RG threshold measured on the CAD test against the number of errors the subject makes on the IT plates. The results reveal the extremely poor correlation between the measured RG color detection thresholds and the number of errors the subjects make on the IT plates within each subject group. Although, in general, subjects with high RG thresholds tend to make more errors, this is not always the case. Within a small range of RG thresholds, both protan and deutan subjects exhibit large differences in IT error rates (Fig. 7a). Both protan and deutan subjects with more than 15 errors on the IT plates exhibit RG thresholds that vary from 3 to 24 SN CAD units.

[Fig. 7 here]

Figure 8 shows the computed plate specific weight (W_i) for each group of color vision subjects. The introductory plate which is read correctly by all subjects has not been included in the computation of SI to ensure that the full range of SI corresponds to a maximum of 24 possible errors. As expected, all plates carry very similar weight for normal trichromats with considerable variation within the two groups of color deficients. The two plates that carry most weight for both deutan and protan subjects are 5 and 21. Interestingly, although significant differences remain, the distribution of weights amongst deutan and protan subjects is very similar. The weights shown in Fig. 8 were used to compute the corresponding SI value for each of 746 subjects investigated and this is plotted in Fig. 7b against the subject's RG threshold. The results are similar to those shown in Fig. 7a and suggest that the large variability in error rates amongst subjects with similar RG thresholds hides any potential benefit the SI index may offer. In order to examine how the mean errors subjects with similar RG thresholds make on the IT plates, the subjects were grouped according to their RG thresholds into bins of width 2.5 CAD SN units. The mean number of errors and the corresponding standard deviations subjects make were then computed within each bin. The procedure was then repeated for the SI values and the results are shown in Fig. 9(a,b).

[Fig. 8 and 9 here]

Discussion

The need to quantify the severity of color vision loss in a simple and effective way has recently become more important, both within occupational environments such as aviation, as well as for clinical applications. An easy solution is to use the number of errors subjects make on color screening tests that employ pseudoisochromatic plates as a measure of severity, but the validity of this approach has not been validated within aviation. The majority of color screening tests are intended to discover congenital color deficiency, but the parameters that emerge from such tests are also used to indicate the severity of color vision loss. The same subjects when tested on different conventional color screening tests can produce quite different results. On some tests the subject can be classified as severe, whilst other tests may only indicate mild loss of chromatic sensitivity. Such inconsistent findings make it difficult to quantify the subject's severity of color vision loss (21). The midpoint and the size of the matching range in anomaloscope matches, for example, are often used to screen for levels of deficiency, even when these parameters do not correlate well with the subject's loss of chromatic sensitivity (4,23). Previous studies also failed to show significant correlation between the number of errors subjects make on IT plates and the parameters of Nagel anomaloscope matches for both color vision deficient (6) and also for normal trichromats (15). In addition, there are other reasons why quantitative, repeatable assessment of color vision has become more important in recent years: the demand from the general public that the methods for occupational testing (of any kind) are transparent; and the requirement that decisions are scientifically based and justifiable by reasons other than "eminent opinion". Also, information concerning occupational testing is more widely available to the public and applicants are more aware of the drawbacks of different tests and of their rights to challenge an adverse decision. A defendable, objective test that can be used to quantify the severity of color vision loss is therefore highly desirable. In addition, one needs to establish pass/fail limits which ensure that applicants with color deficiency that pass within aviation or other occupational environments can carry out visually-demanding, color-related tasks with the same accuracy as normal trichromats.

In this paper we also present data and analysis of RG and YB color detection thresholds measured under conditions which isolate the use of color signals. The thresholds exhibit an almost linear relationship with the corresponding cone contrasts generated by the colored stimuli, even for CDs that are 10 times larger than the median threshold values for normal trichromats. Departures from linearity are observed for much larger threshold values and these become particularly significant for S-cone activation and less so for L- and M-cones (Fig. 3). The measure of CD employed to quantify the subject's color thresholds correlates well with the luminance contrast of a similar achromatic stimulus that matches the perceived contrast of the colored stimulus (Fig. 4). These observations suggest that RG and YB color detection thresholds when measured in the CIE 1931- (x,y) chromaticity chart can be used to describe appropriately the severity of color vision loss.

The main thrust of this investigation was to examine the extent to which the use of the number of errors subjects make on the IT plates as a measure of the severity of RG loss can be improved by producing appropriate weights that describe the probability of a correct response for each plate. Fig. 5 shows the probability of making, *k*, or less errors on the IT plates when the maximum possible number of errors is 24. These data show clearly that one cannot treat deutan and protan subjects that make the same number of errors as equivalent in terms of their loss of chromatic sensitivity. No such distinction is currently made within aviation. The results also show that only 80.9% of normal trichromats pass with zero errors and that four or less errors must be allowed for to ensure that no normal trichromat is disadvantaged. Even if only three or less errors are allowed, 10% of deutan and 1% of protan subjects also pass. The distribution of errors subjects make are plate-specific (see Fig. 7) and this leads to different weights being produced for each plate within each subject group (Fig. 8). The current practice of allowing varying numbers of errors as a pass depending on the color-related visual demands within a given occupation does not take into account either the class of deficiency involved or the difficulty of the plates failed.

The number of errors subjects make on the IT plates relates non-linearly to the corresponding RG thresholds for both deutan and protan subjects. The large variability observed in error rates, even for subjects with very similar RG thresholds reduces the usefulness of this parameter as an indicator of the severity of color vision loss. Fig. 8 shows clearly that the probability of a correct response varies significantly across the 24 plates and that the observed variation is different for each of the three groups. The computed SI takes into account the probability of a correct response when the subject misreads the number on a given plate and therefore it is reasonable to expect that the SI value would be more appropriate to describe the severity of color vision loss. Although the SI value may well provide some improvement (Fig. 7b), this is difficult to assess since the inter-subject variability remains high. By examining average IT errors made by subjects with similar RG thresholds (Fig. 9), the results for both the number of errors and the SI values reveal more clearly the non-linear relationship when plotted against the subject’s RG threshold. Initially both the number of errors and the SI values increase rapidly with RG threshold. For RG thresholds greater than ~ 10 SN units, neither the number of errors nor the SI values reflect well the further loss of chromatic sensitivity, particularly for deutan subjects.

The spread in the number of errors and SI values observed in normal trichromats is of interest and can be accounted for mostly in terms of variability in their chromatic sensitivity. There is little doubt that a strong chromatic signal contributes most to the definition of the contours that delineate the numerals employed on IT plates. An examination of the spread in RG color thresholds within normal trichromats reveals a 2.2 fold variation in RG chromatic sensitivity. Although this variation is small, subjects with low thresholds will have an advantage over subjects at the extreme upper end of the normal threshold range. When coupled with other factors such as cultural differences, age, the use of serif fonts for number representation, concentration, factors that are equally applicable to color deficients, etc., normal trichromats with reduced chromatic sensitivity may well perform less well and produce the error scores observed experimentally.

Conclusions

1. The magnitude of RG and YB color thresholds when expressed as a displacement in the CIE 1931 –(x,y) chromaticity chart and referenced to background chromaticity relate almost linearly to the corresponding cone contrasts generated by the colored stimulus. The perceived contrast /

conspicuity of suprathreshold, colored stimuli relates linearly to the luminance contrast of a spatially similar, achromatic stimulus. These findings suggest that the magnitude of color detection thresholds measured in this way may be a good descriptor of the severity of color vision loss in subjects with congenital deficiency.

2. The first 25 plates of the Ishihara test (38 plate ed.) have different probabilities of eliciting a correct response. The ranking order of increasing probability is group specific and therefore the same number of errors made on the IT plates by subjects from different groups do not reflect the same severity of color vision loss. The number of errors the applicant makes on the IT plates should not therefore be used by occupational medical advisors to judge suitability for the job.

3. The computation of a severity index that takes into account the plate-specific probability of a correct response within each group has only marginal benefit because of the large variability in SI values measured in subjects with similar threshold chromatic sensitivity. Although on average subjects with large RG thresholds will have higher SI values, neither the number of errors made on IT plates nor the SI value computed from these errors is a guaranteed indicator of the subject's loss of chromatic sensitivity.

Reference List

[1] Ahnelt PK, Kolb H, Pflug R. Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. J Comp Neurol 1987; 255(1):18-34.

[2] Barbur JL. 'Double-blindsight' revealed through the processing of color and luminance contrast defined motion signals. Prog Brain Res 2004; 144:243-259.

[3] Barbur JL, Harlow AJ, Plant GT. Insights into the different exploits of color in the visual cortex. Proc R Soc Lond B Biol Sci 1994; 258(1353):327-334.

[4] Barbur JL, Rodriguez-Carmona M, Harlow JA, Mancuso K, Neitz J, Neitz M. A study of unusual Rayleigh matches in deutan deficiency. Vis Neurosci 2008; 25(3):507-516.

[5] Belcher SJ, Greenshields KW, Wright WD. Color vision survey using the Ishihara, Dvorine, Bostrom and Kugelberg, Bostrom, and American-Optical Hardy-Rand-Rittler tests. Br J Ophthalmol 1958; 42(6):355-359.

[6] Birch J. Efficiency of the Ishihara test for identifying red-green color deficiency. Ophthalmic Physiol Opt 1997; 17(5):403-408.

[7] Birch J, McKeever LM. Survey of the accuracy of new pseudoisochromatic plates. Ophthalmic Physiol Opt 1993; 13(1):35-40.

[8] Civil Aviation Authority. Minimum colour vision requirements for professional flight crew - Part III Recommendations for new colour vision standards. The Stationary Office; May 2009. Report No: CAA Paper 2009/04.

[9] Cosstick M, Robaei D, Rose K, Rochtchina E, Mitchell P. Numerical confusion errors in ishihara testing: findings from a population-based study. Am J Ophthalmol 2005; 140(1):154-156.

[10] Federal Aviation Administration. Version V of the guide for aviation medical examiners. Retrieved May 19, 2011, from: http://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/ame/guide/app_process/exam_tech/item52/amd/

[11] Frey RG. Most suitable pseudoisochromatic table for practice. Comparative studies on applicability of Bostrom, Bostrom-Kugelber, Hardy-Rand-Ritter, Ishihara, Rabkin and Stilling tables. Albrecht Von Graefes Arch Ophthalmol 1958; 160(3):301-320.

[12] Haskett MK, Hovis JK. Comparison of the Standard Pseudoisochromatic Plates to the Ishihara color vision test. Am J Optom Physiol Opt 1987; 64(3):211-216.

[13] Jennings BJ, Barbur JL. Color detection thresholds as a function of chromatic adaptation and light level. Ophthalmic Physiol Opt 2010; 30(5):560-567.

[14] Joint Aviation Authorities. JAR-FCL 3. Flight Crew Licensing (Medical). Hoofddorp, The Netherlands: JAA; 2002. Section 1, subpart B. Class 1 Medical Requirements (amendment 2).

[15] Miyahara E. Errors reading the Ishihara pseudoisochromatic plates made by observers with normal color vision. Clin Exp Optom 2008; 91(2):161-165.

[16] Perales J., Hita E. Influence of some factors on non-typical responses to three tests of color vision in children. In: Verriest G., editor. Color Vision Deficiencies VII. The Hague: Junk, 1984: 211-219.

[17] Pokorny J, Smith VC, Verriest G, Pinckers AJLG. Congenital and Acquired Color Vision Defects. New York: Grune & Stratton, 1979.

[18] Rodriguez-Carmona M. Variability of chromatic sensitivity: fundamental studies and clinical applications [PhD Thesis]. London, United Kingdom, City University; 2006.

[19] Rodriguez-Carmona M, Harlow AJ, Walker G, Barbur JL. The Variability of Normal Trichromatic Vision and the Establishment of the 'Normal' Range. Proceedings of 10th Congress of the International Color Association, Granada (Granada, 2005) 2005;979-982.

[20] Sloan LL, Habel A. Tests for color deficiency based on the pseudoisochromatic principle; a comparative study of several new tests. AMA Arch Ophthalmol 1956; 55(2):229-239.

[21] Squire TJ, Rodriguez-Carmona M, Evans ADB, Barbur JL. Color vision tests for aviation: comparison of the anomaloscope and three lantern types. Aviat Space Environ Med 2005; 76:421-429.

[22] Walkey HC, J.L.Barbur, Harlow A, Makous W. Measurements of Chromatic Sensitivity in the mesopic range. Color research and application 2001; supplement vol 26:S36-S42.

[23] Wright WD. Researches on Normal and Defective Color Vision. London: Henry Kimpton, 1946.

Table I: Number of subjects that pass the Ishihara Test for different accepted testing protocols within various professional environments.

Professional environment:		JAR†	FAA†			London Underground Ltd.	Fire Service UK
Edition:	38-plate ed	24-plate ed	38-plate ed	24-plate ed	14-concise	38-plate ed	24-plate ed
Pass Criteria:	1-25 plates no errors	1-15 plates no errors	1-21 plates <=8 errors	1-15 plates <=6 errors	1-10 plates <=5 errors	1-17 plates no errors & <=3 esp‡	1-17 plates <=2 errors
N* (236)	191	213	236	236	236	229	235
D* (340)	2	5	68	64	108	5	24
P* (166)	0	0	10	10	14	1	2
% N*	80.93	90.25	100.00	100.00	100.00	97.03	99.58
% D*	0.59	1.47	20.00	18.82	31.76	1.47	7.06
% P*	0.00	0.00	6.02	6.02	8.43	0.60	1.20

*Color vision class: N=normal; D=deutan; P=protan; †see main text for definition; ‡esp = errors on specified plates.

Table II: Summary of the errors made by some of the 236 normal trichromats on the Ishihara plates 1-25 (38-plate edition).

Plate No.	Correct reading	Colour defective reading	Number read and incidence in normal trichromats (n)	No. errors (n)	n as % of total
1	12	12		0	0.0
2	8	3		0	0.0
3	6	5		0	0.0
4	29	70		0	0.0
5	57	35		0	0.0
6	5	2		0	0.0
7	3	5	8 (5)	5	2.1
8	15	17	19 (1)	1	0.4
9	74	21	71 (2), 24 (1)	3	1.3
10	2	-	8 (1)	1	0.4
11	6	-	8 (1)	1	0.4
12	97	-	87 (24), 37 (1)	25	10.6
13	45	-		0	0.0
14	5	-		0	0.0
15	7	-		0	0.0
16	16	-	46 (1)	1	0.4
17	73	-	78 (6), 23 (4), 13 (2), 77 (1)	13	5.5
18	-	5	17 (1), 47 (1)	2	0.8
19	-	2	**2 (5)**, 8 (2)	7	3.0
20	-	45	47 (1)	1	0.4
21	-	73		0	0.0
22	26	6 or 2	28 (2)	2	0.8
23	42	2 or 4	45 (1)	1	0.4
24	35	5 or 3		0	0.0
25	96	6 or 9		0	0.0

Table and Figure legends

Table I: Number of subjects that pass the Ishihara Test for different accepted testing protocols within various professional environments.

Table II: Summary of the errors made by some of the 236 normal trichromats on the Ishihara plates 1-25 (38-plate edition).

Figure 1: Steps involved in the computation of cone-contrast along any line in (x,y) – CIE 1931 chromaticity space, as a function of chromatic displacement away from background chromaticity. The approach described here (see methods section) allows the computation of contrast for any class of photoreceptors without making any other assumptions.

Figure 2: Cone contrasts computed for chromatic displacements defined by the mean ellipse shown in Fig. 1a. A hue angle of 0^{o} corresponds to a horizontal direction towards the long wavelength region of the spectrum locus.. The cone contrast diagram (enlarged in b) shows clearly the directions of chromatic displacement that correspond to ~ zero L- and M-cone contrasts: ~ 67^{o} and 247^{o} and also the two directions for which the S-cone contrast is zero: 154^{o} and 334^{o} .

Figure 3: Cone contrasts generated along the YB-axis (67^{o} and 247^{o}) and the RG-axis (154^{o} and 334^{o}). The graphs show maximum chromatic displacement distance (CD) of ~ 14 standard normal (SN) thresholds units for YB discrimination and ~ 23 SN threshold units for RG discrimination. The non-linearity increases with the size of CD involved, particularly for the YB axis, but the departure from linearity is small for both RG and YB thresholds less than ~ 10 SN units.

Figure 4: The perceived contrast of a photopically isoluminant colored stimulus (see inset) is matched with that of an 'equivalent' achromatic stimulus. The stimuli were presented for 0.5 secs and the subject's task was to indicate which of the two stimuli had the highest perceived contrast (or

conspicuity). A two alternative staircase procedure was used to estimate the contrast of the achromatic stimulus that matched the perceived contrast of a colored stimulus for each value of CD investigated.

Figure 5: The probability of making (*k*) or less errors when reading the numerals on the Ishihara test plates plotted for a group of normal trichromats and for subjects with congenital, deutan and protan color deficiency. The order of presentation was random for the remaining 24 plates of the Ishihara 38-plate test.

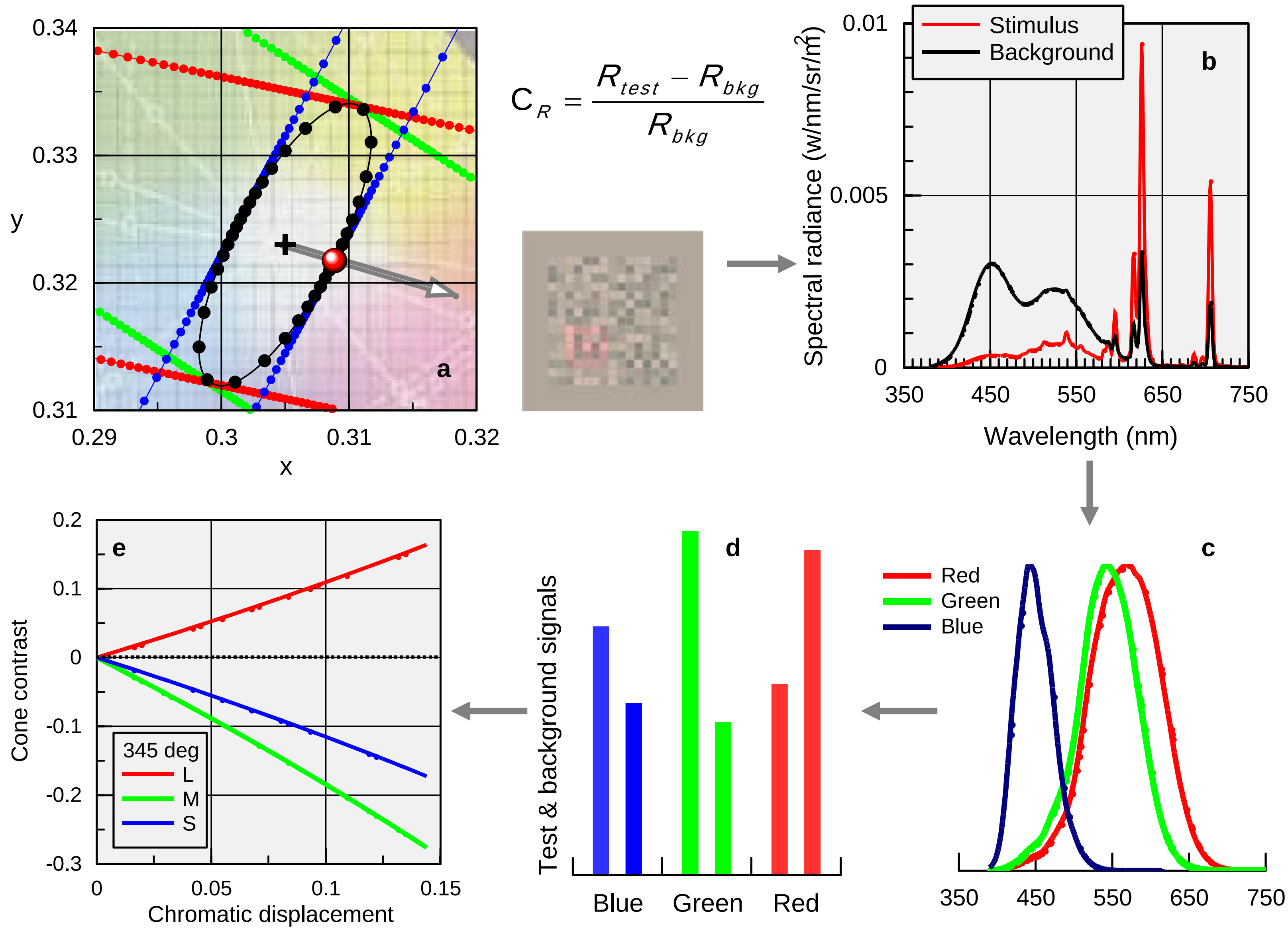
Figure 6: The percentage of subjects in each group that make errors on each of the 25 plates plotted in a ranked sequence. The numbers along the horizontal axis indicate the plate number in the ranked sequence. Note that these numbers are different depending on the class of color vision involved.

Figure 7: (a) The number of errors made by each subject on the Ishihara test plotted against the subject's RG threshold (measured in SN CAD units). (b) Graph showing the 'Severity Index' (SI) plotted against the subject's CAD RG threshold. The SI varies from zero (i.e., no errors) to 100 (when the subject makes errors on every one of the 24 plates). The SI takes into account both the number and difficulty of the plates failed. Data are shown for normal trichromats (grey diamonds), deutan (green discs) and protan (red squares) subjects

Figure 8: Graph showing the 'weight' of each individual Ishihara plate (38-plate edition) for each of the three subject groups. The 'weight' assigned to each plate is the probability of making no errors scaled appropriately so to ensure that the sum of the weights for the 24 plates employed is 100. Plates with small weights are more difficult to read correctly than plates with large weights. If a subject makes an error on a plate with a large weight (i.e., a plate that most subjects can read correctly), this error indicates a greater loss of chromatic sensitivity.

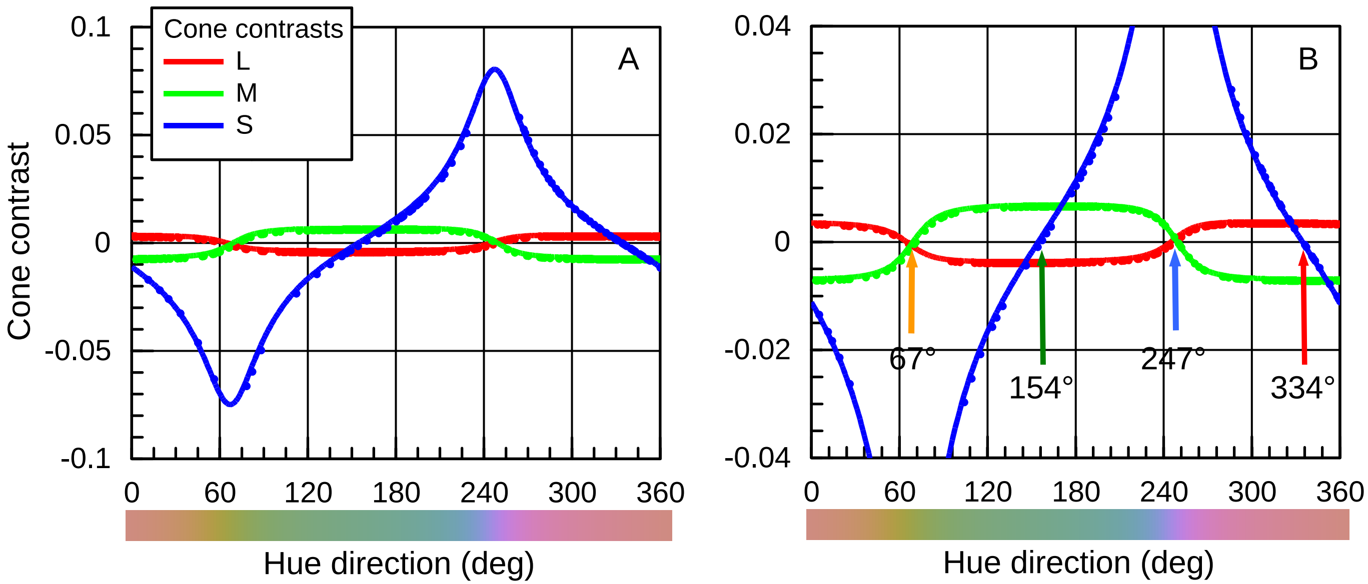
Figure 9: The spread in mean error scores (a) and the computed SI of color loss (b) are plotted as a function of CAD RG threshold for bin widths of 2.5 SN CAD units. The error bars show ±2 standard deviations from the computed mean.

Figure 1



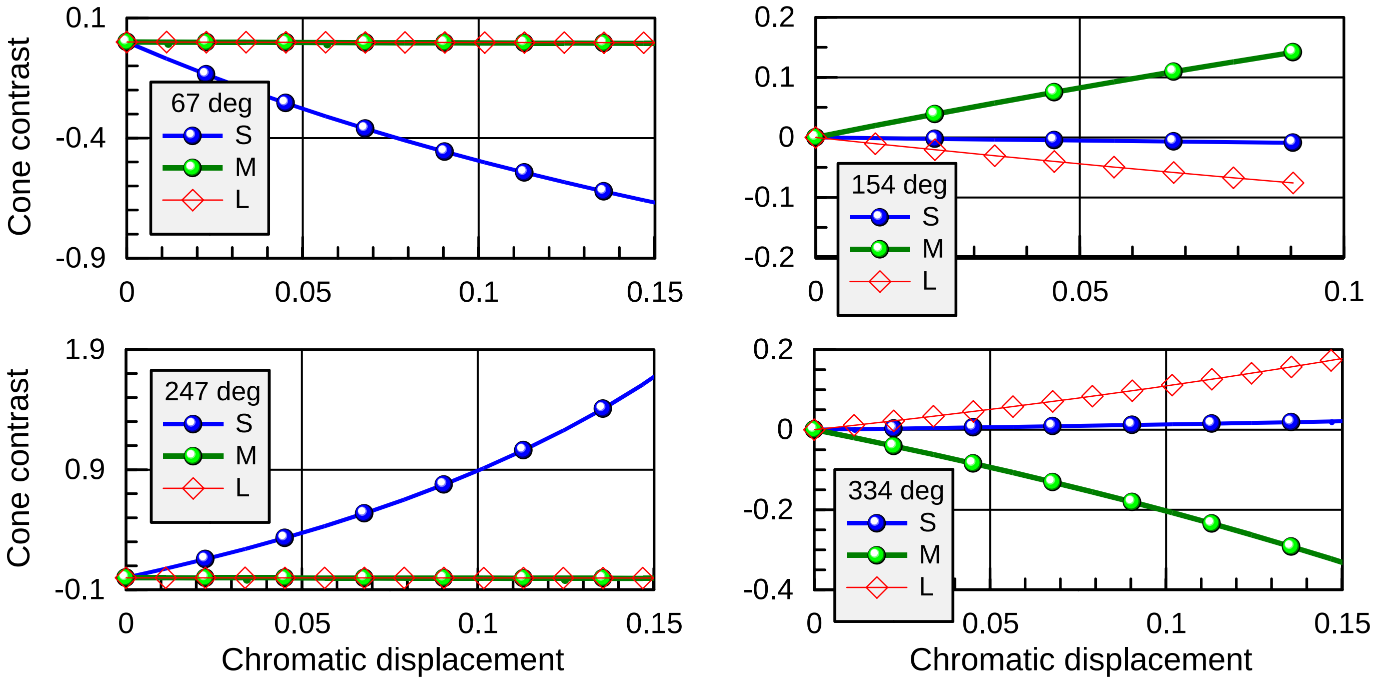
0.34
0.33
0.32
0.31
y
0.29
0.3
0.31
0.32
x
a
$$C_R = \frac{R_{test} - R_{bkg}}{R_{bkg}}$$
Stimulus
Background
b
0.01
0.005
0
Spectral radiance (w/nm/sr/m^2)
350
450
550
650
750
Wavelength (nm)
c
Red
Green
Blue
350
450
550
650
750
d
Test & background signals
Blue
Green
Red
e
0.2
0.1
0
-0.1
-0.2
-0.3
Cone contrast
345 deg
L
M
S
0
0.05
0.1
0.15
Chromatic displacement

Figure 2



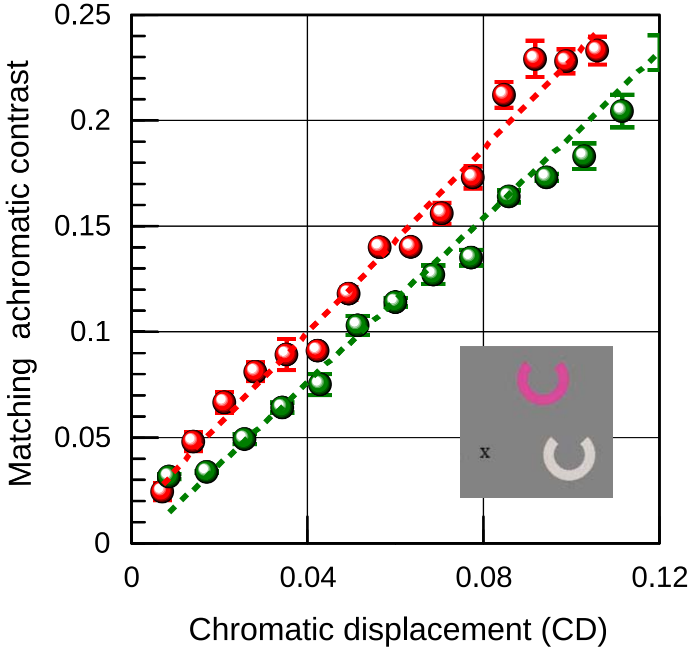
Cone contrasts
L
M
S
A
B
0.1
0.05
0
-0.05
-0.1
0.04
0.02
-0.02
-0.04
0
60
120
180
240
300
360
Cone contrast
Hue direction (deg)
67°
154°
247°
334°

Figure 3



0.1
-0.4
-0.9
Cone contrast
67 deg
S
M
L
0
0.05
0.1
0.15
0.2
0.1
0
-0.1
-0.2
154 deg
S
M
L
0
0.05
0.1
1.9
0.9
-0.1
Cone contrast
247 deg
S
M
L
0
0.05
0.1
0.15
Chromatic displacement
0.2
0
-0.2
-0.4
334 deg
S
M
L
0
0.05
0.1
0.15
Chromatic displacement

Figure 4



0.25
0.2
0.15
0.1
0.05
0
Matching achromatic contrast
0
0.04
0.08
0.12
Chromatic displacement (CD)
x

Figure 5

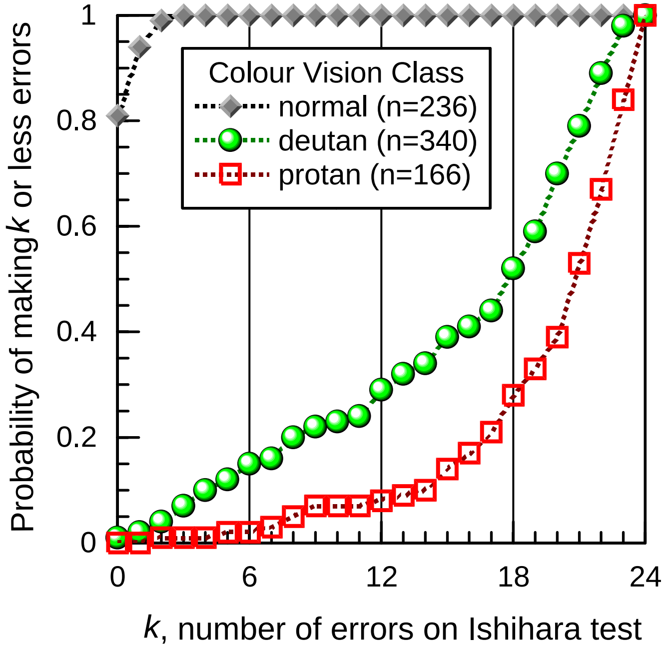
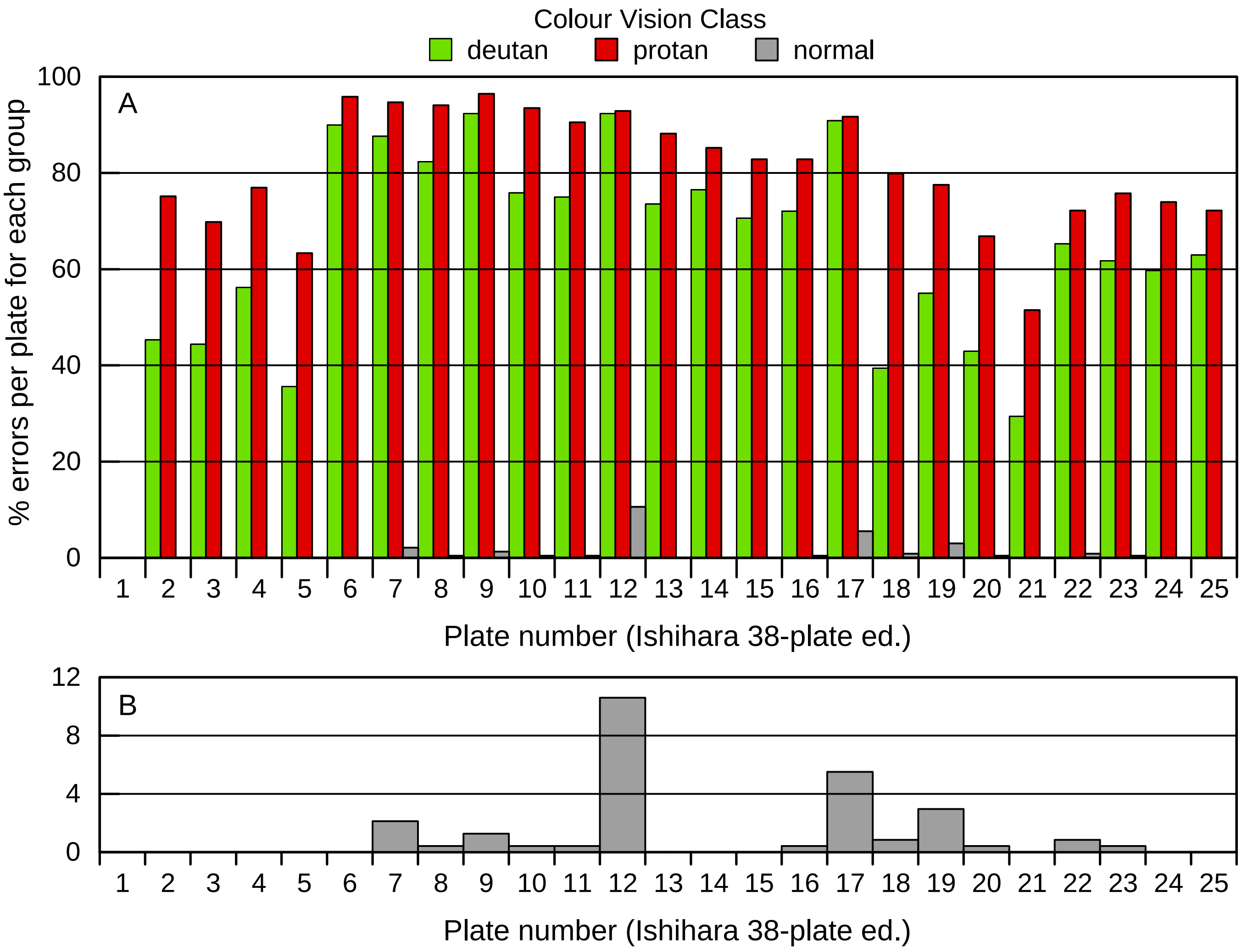
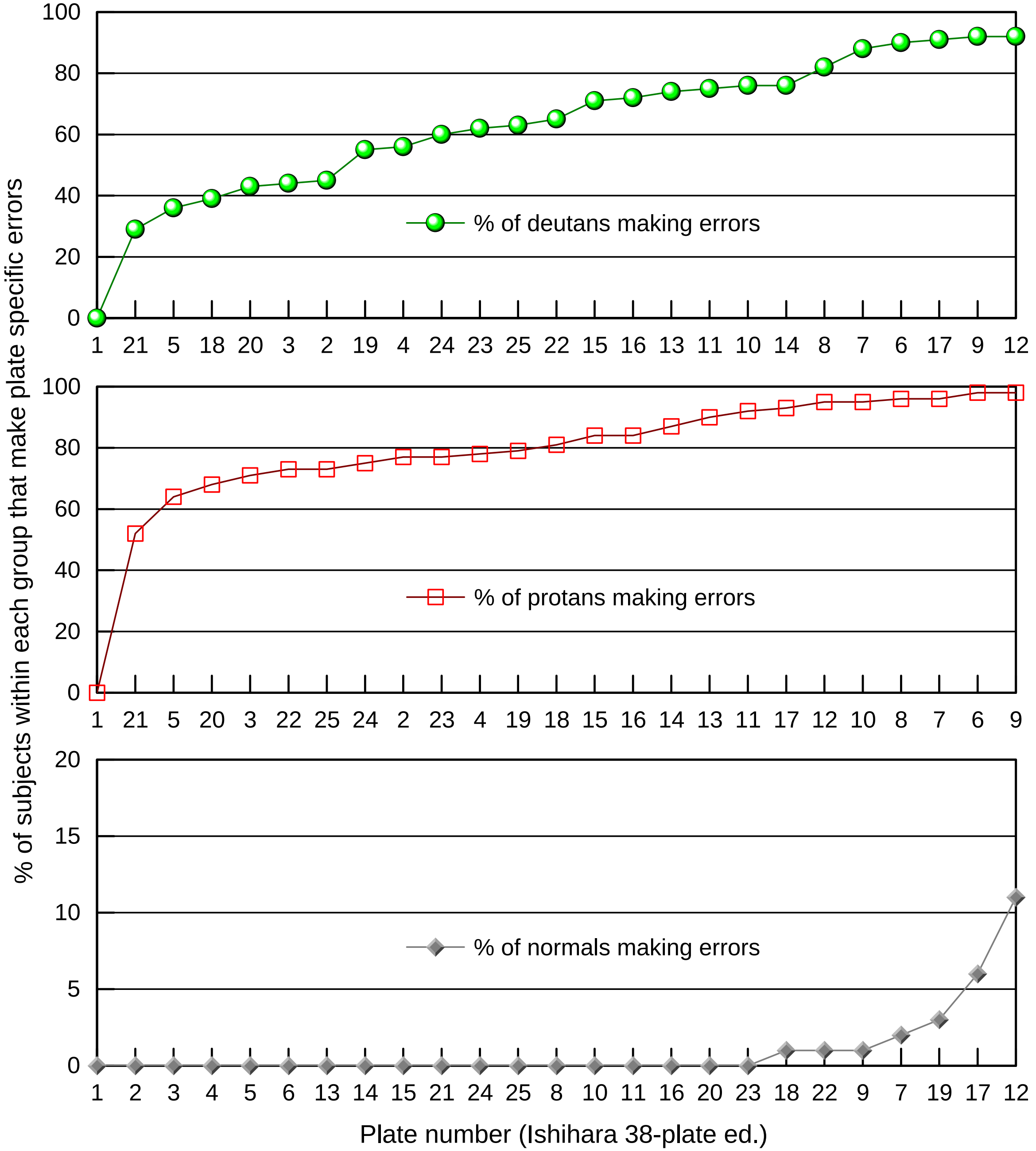

Colour Vision Class
normal (n=236)
deutan (n=340)
protan (n=166)
Probability of making k or less errors
k, number of errors on Ishihara test
0
0.2
0.4
0.6
0.8
1
0
6
12
18
24

Figure 6



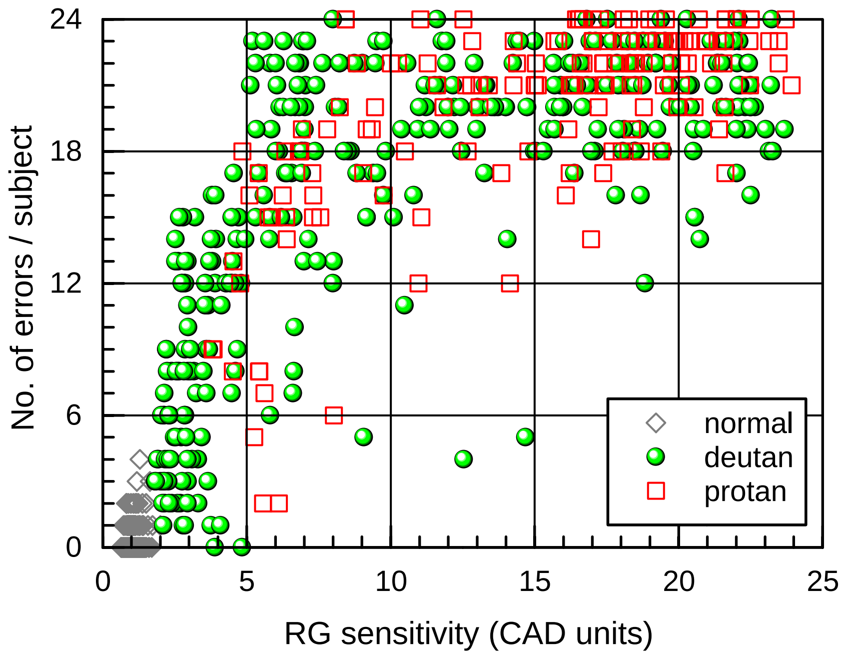
Colour Vision Class
deutan
protan
normal
100
80
60
40
20
0
% errors per plate for each group
A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Plate number (Ishihara 38-plate ed.)
12
8
4
0
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Plate number (Ishihara 38-plate ed.)

Figure 7



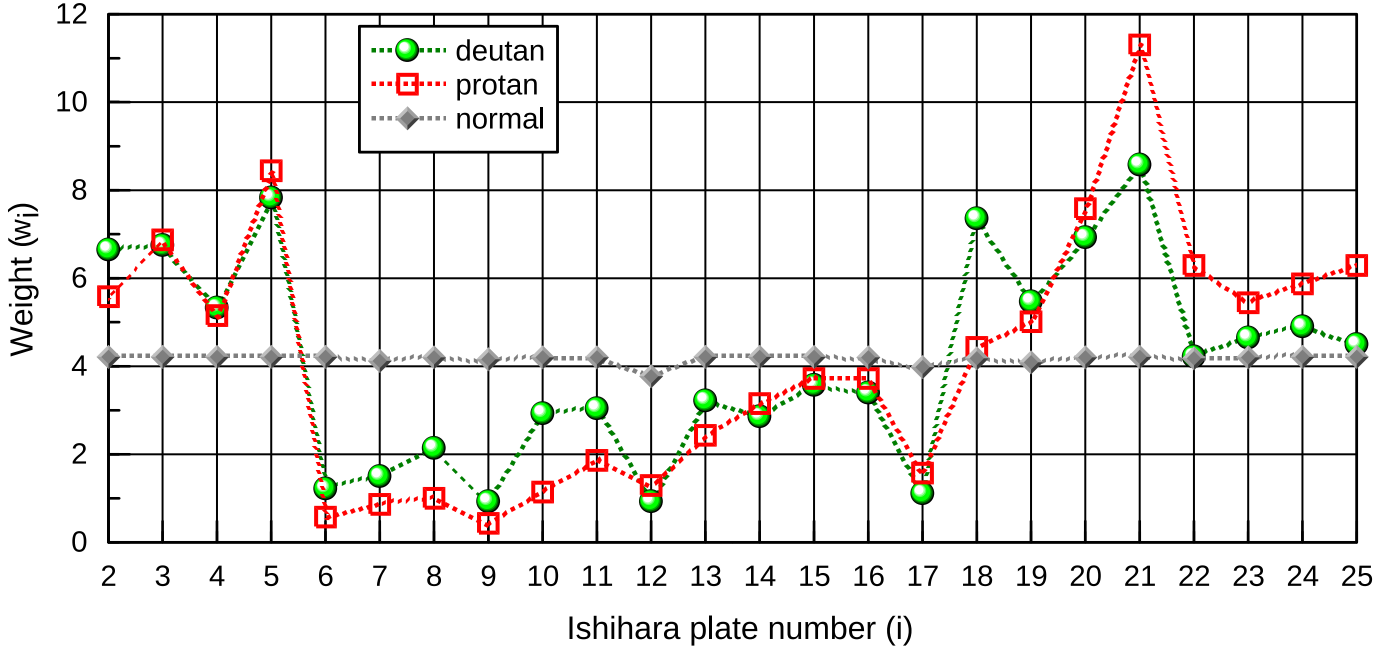
% of subjects within each group that make plate specific errors
% of deutans making errors
1 21 5 18 20 3 2 19 4 24 23 25 22 15 16 13 11 10 14 8 7 6 17 9 12
% of protans making errors
1 21 5 20 3 22 25 24 2 23 4 19 18 15 16 14 13 11 17 12 10 8 7 6 9
% of normals making errors
1 2 3 4 5 6 13 14 15 21 24 25 8 10 11 16 20 23 18 22 9 7 19 17 12
Plate number (Ishihara 38-plate ed.)

Figure 8



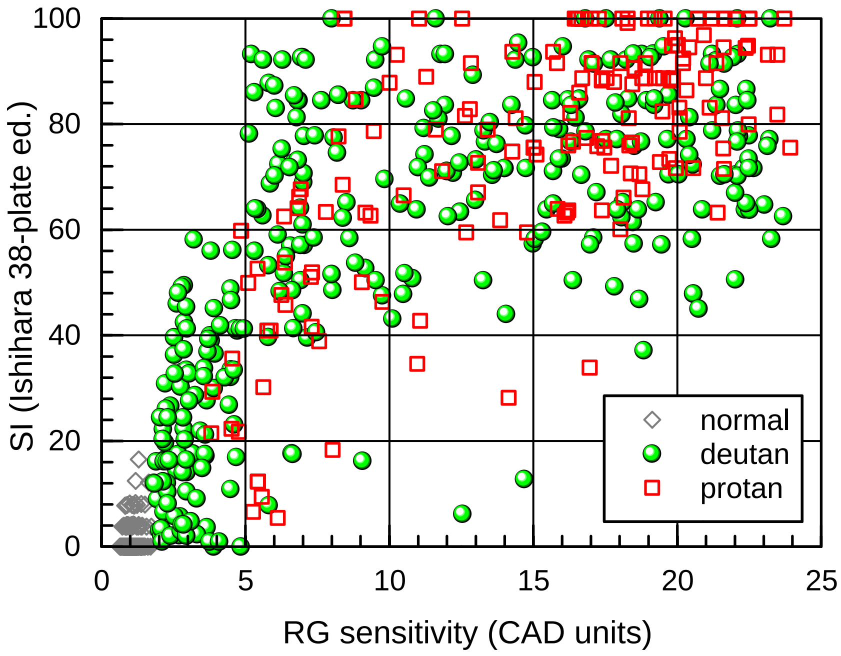
No. of errors / subject
24
18
12
6
0
0
5
10
15
20
25
RG sensitivity (CAD units)
normal
deutan
protan

Figure 9



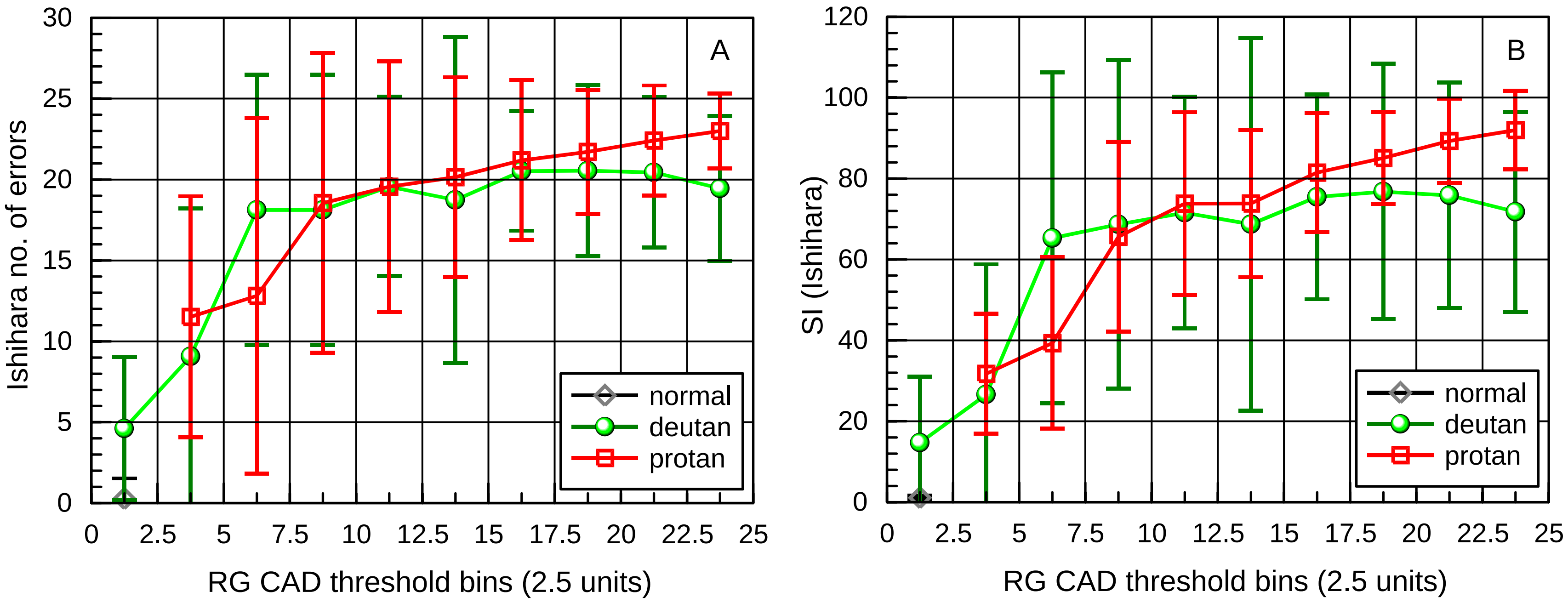
deutan
protan
normal
Weight (w_i)
Ishihara plate number (i)
0
2
4
6
8
10
12
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figure 10



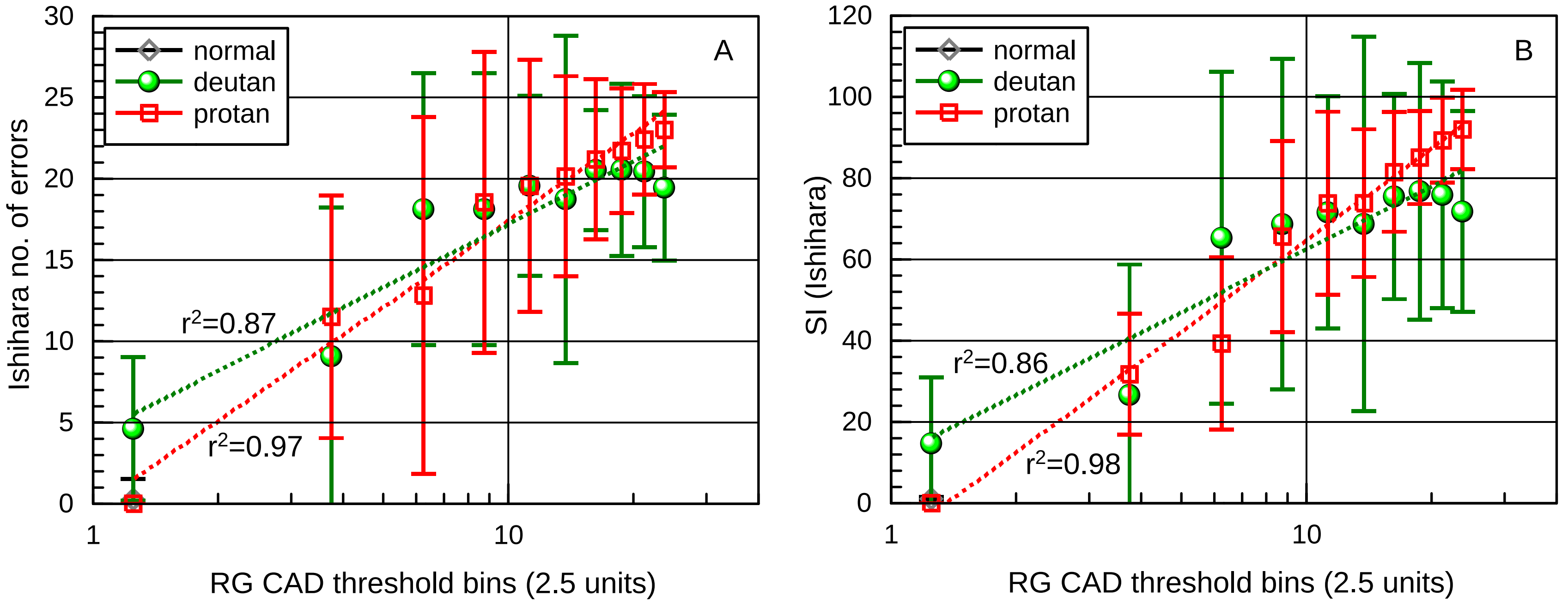
100
80
60
40
20
0
0
5
10
15
20
25
SI (Ishihara 38-plate ed.)
RG sensitivity (CAD units)
normal
deutan
protan

Figure 11



A
B
Ishihara no. of errors
SI (Ishihara)
normal
deutan
protan
normal
deutan
protan
RG CAD threshold bins (2.5 units)
RG CAD threshold bins (2.5 units)

Figure 12



A
normal
deutan
protan
Ishihara no. of errors
r2=0.87
r2=0.97
RG CAD threshold bins (2.5 units)
B
SI (Ishihara)
r2=0.86
r2=0.98
RG CAD threshold bins (2.5 units)